Award No. G12AP20137

National Geological and Geophysical Data Preservation Program

Washington State Metadata Project:

Washington Industrial Minerals Metadata Production

Final Technical Report

David Jeschke/Robert Berwick/Stefanie Kinnamon

Washington Department of Natural Resources

Division of Geology and Earth Resources

PO Box 47007

Olympia, WA 98504-7007

(360) 902-1495 / (360) 902-1450

(360) 902-1785 fax

david.jeschke@dnr.wa.gov

robert.berwick@dnr.wa.gov

stefanie.kinnamon@dnr.wa.gov

Award Term: July 1, 2012 to June 30, 2013

Submitted: June 28, 2013

Abstract

The Washington State Department of Natural Resources, Division of Geology and Earth Resources, houses many geologic and geophysical data collections, 16 of which have been inventoried in the National Digital Catalog. For our 2012 project for the National Geological and Geophysical Data Preservation Program (NGGDPP), we focused on our industrial minerals publications, files and spreadsheets.

Our objective was to create site specific metadata for submission to the National Digital Catalog and consolidate existing databases and spreadsheets into a single downloadable geospatial relational database that would include hyperlinks to scanned documents, reports and maps that were previously only available by physically accessing the paper files.

To accomplish this, our staff scanned our metal mine files and flat map collection covering 1,902 industrial mineral occurrences. We entered data for 1,252 scanned pages and compiled databases and spreadsheets into a single database which contains 4,273 hyperlinks to scanned documents and databases. This information was compiled into XML format in accordance with the National Digital Catalog specifications, and submitted for inclusion in the catalog.

Table of Contents

Introduction

The Washington State Department of Natural Resources, Division of Geology and Earth Resources (DGER), identified 16 collections of geologic and geophysical data which have been inventoried in the National Digital Catalog.

In 2009, the first year we participated in the National Geological and Geophysical Data Preservation Program (NGGDPP), we identified three collections that contain a wealth of high-value geologic and geophysical information that were at imminent risk of degradation or loss, either through physical deterioration, lack of documentation, or disposal. These collections were rock core, rock cuttings and geotechnical reports. We produced XML metadata records for 47 boreholes that produced rock cores, 309 boreholes that produced rock cuttings, and 56,302 boreholes represented in 10,000 geotechnical reports for a total of 56,658 total metadata records submitted to the National Digital Catalog.

In 2010 DGER continued to add to the geotechnical report collection. During this year we provided National Digital Catalog with an additional 28,196 records which were derived from 8,053 reports. The Division of Geology and Earth Resources also launched the Washington State Geologic Information Portal, an interactive mapping site. This portal was designed to make data available and accessible which addresses one of the core requirements of the Geologic Data Preservation Plan of Washington.

In 2011 DGER began work preserving mining data from our paper and electronic files. We scanned our metal mine files and flat map collection which covered 4,050 mining sites. In addition, we entered data for 29,436 scanned pages and compiled databases and spreadsheets into a single database containing 106,246 hyperlinks to scanned documents and databases.

When the Washington Geological Survey was first established, one of the main directives was to collect drawings, maps, reports, minerals, and other information relating to the mineral industry and make this information accessible to the public. We continue this goal by selecting the industrial minerals files for our 2012 data preservation project.

The information contained in these mining files is invaluable. Since 1890 the Washington Geological Survey has amassed an extensive collection of historical industrial mineral-related data on over 1,900 sites that include such information as ownership, mine maps, reports, assays, geochemical surveys, and production records. These files are unique, and are often the only source of information about mineral occurrences, locations, production, ownership and the type of material present.

The importance of this information is underscored by the diversity of ways in which it is used, and the people who are using it. Not only is this information currently being used by the mining industry in one of the most extensive mineral exploration programs that Washington has seen in years, it is also being used in litigation over mine contamination in the upper Columbia River and other areas of the state. The collection is routinely used by local, state, and federal agencies, lawyers, small miners, mining companies, researchers, mining and engineering students, realtors, educators, and the general public.

Project Goals

The purpose of this project was to preserve physical data by converting paper files to digital format, and consolidate these data with existing industrial mineral-themed spreadsheets and documents. The plan was to consolidate into a relational database, create FGDC compliant metadata, and generate XML metadata for submittal to the National Digital Catalogue.

At the beginning of the project DGER estimated that the number of items for which we would evaluate for scanning and create metadata for approximately 8,000 pages of historical documents related to approximately 1,700 mineral occurrences.

The project continued over a one year period beginning July 1, 2012 and ending June 30, 2013. The initial plan was to begin by designing a schema to contain information about each industrial mineral occurrence. Over the next five months the relevant publications would be identified and attribute information about each occurrence entered into the schema.

During the following five months point locations were to be identified for each of the mineral occurrences. A location confidence code would be produced for each occurrence to describe the quality of the available location information.

The last 6 weeks we reserved for the data consolidation, and to complete FGDC metadata for the feature class and the related table in the geodatabase. Also during this time we would compile and submit the XML metadata to the National Digital Catalogue.

Methodology

We began our project by building a timeline to track progress. This timeline included important milestones to be achieved which staff would be working on that portion of the project and estimates of the amount of time that would be required. This timeline was referred to frequently during the project to insure that deliverables would be produced on

schedule.

A schema was built in Excel for the entry of all attribute information about each of the industrial mineral occurrences in Washington Division of Geology and Earth Resources (DGER) Bulletin 37. Attribution for 1,693 occurrences was entered into this schema.

Meanwhile we conducted a literature search for other DGER and United States Geological Survey (USGS) publications which contained industrial mineral occurrences in Washington State. Several publications were identified including: DGER Bulletins 48, 49, 51 & 52, and USGS Bulletin 1402.

Each of the 1,693 occurrences identified from Bulletin 37 was geolocated and entered into a feature class using ArcGIS Desktop. The location described in Bulletin 37 was used to find the closest possible point at which to locate the occurrence. Additional sources including web searches, Google Maps and Google Earth were also utilized to identify the most correct possible location. For each point geolocated a point identifier was generated to link the point back to the correct attribution record in the Excel spreadsheet. A location confidence code was entered for each point location. This code indicates how precisely the available information allowed us to locate the point.

After this, we reviewed approximately 8,000 pages of documents from DGER files. We determined which of these documents to scan for inclusion in the final product based on: content regarding the quality, location or value of the occurrence. Documents which were redundant or added no useful information were excluded. A total of 1,252 pages were identified, scanned, and linked in the deliverable. We built a table in Excel to contain attribution such as the author, title and publication date for each of the scanned documents.

The other publications which contain industrial mineral occurrences were reviewed. It was important that this step occur after all the Bulletin 37 points were geolocated so that redundant entries would be easier to identify. An additional 209 occurrences were located and geocoded from these publications and attribution for each of these occurrences was entered as the points were located.

The Excel spreadsheet containing information about the scanned documents was imported into an ESRI file geodatabase. Links were built between the point location feature class and this document table. A total of 4,273 links were constructed. The feature class and document table were error checked for quality assurance. FGDC compliant metadata was written for both the feature class and document table.

We moved the scanned files to a designated storage space used to host files where the public can access them, and created hyperlinks to the scanned documents. A downloadable database package will to be placed on our website containing the industrial mineral locations and the

related document table. We generated XML metadata suitable for inclusion in the National Digital Catalog. Metadata was submitted to Natalie Latysh on June 26, 2013.

Results

We submitted XML metadata for 1,902 sites that includes links to 1,252 pages of scanned documents for the grant period July 1, 2011 to June 30, 2012. The delivery date of the metadata to the contracting official was June 26th and the industrial minerals database was made available for download by June 28th, both of which our contract requirements.

Below is a comparison of our project goals to our actual accomplishments during the grant period (July 1, 2011 to June 30, 2012).

Estimated	Number of industrial mineral occurrences	~1,700
	Number of pages to review for scanning	~8,000
Results	Number of industrial mineral occurrences	1,902
	Number of pages scanned	1,252

We estimated that we would find records for approximately 1,700 industrial mineral occurrences within Washington State. After a thorough review of all the relevant publications we found a total of 1,902 occurrences, all of which are documented in our database.

We estimated that there were approximately 8,000 pages to review in the industrial minerals files. As anticipated, many of these pages turned out to be redundant or irrelevant and were not scanned. A total of 1,252 pages of scans were completed and linked to the mineral occurrences.

In the future we plan to create a mining theme, including the metal mines and industrial mineral occurrences, in the Washington State Geologic Portal, our interactive mapping application:

http://www.dnr.wa.gov/ResearchScience/Topics/GeosciencesData/Pages/geology_portal.aspx

The Washington State Geologic Information Portal utilizes the ArcGIS Server interactive mapping application and is programmed with the Flex API, which is based on the Adobe Flash platform. One of the many strengths of ArcGIS Server with Flex is serving large datasets to the user quickly. With this application the user can navigate through different themes; locate, overlay, and query data; create maps with titles, legends, and labels; create their own points and polygons; and export their maps as PDFs or geo-referenced TIFFs.